

CSV Conversion to IATI

(draft for feedback and comments)

Product vision

Context

Producing an IATI file is (partly) a technical challenge: data may be available in existing information systems or spreadsheets, and needs to be transformed into XML data. This challenge has three components:

- **Combining** data from various sources: finance, content, results information may be kept in separate systems.
- **Converting** the data into IATI XML and getting feedback on compliance and quality of the data.
- **Publishing** the data by making it available online and (if needed) updating the IATI Registry.

Some organisations may use integrated information systems that can export IATI data by themselves: the tool solves both of the first two challenges, and possibly the third as well.

Organisations with a limited number of activities and transactions could copy-paste or upload their information into AidStream¹. They could then download all their information as either XML or as a spreadsheet. Adding and updating information still needs to happen in the AidStream interface.

- Combining the data is done by hand.
- Converting and publishing the data is done by AidStream.

All other organisations will need to set up their own tools and processes to address the three challenges. We propose a new “combine-convert” tool or service, to address the first two challenges².

It may be possible to include a component to publish the data, but this is less of an ongoing challenge for IATI officers.

There is an additional conversion step to mention: IATI uses many codelists and vocabularies, for instance for sector classifications, types of transactions, and so on. In version 2 of IATI, an effort has been made to make the codes themselves language-neutral, but as a consequence, they are not easily human-readable. A translation or conversion from

¹ A similar approach is also possible with Akvo RSR.

² See the appendix at the end for a list of tools and services that (partially) try to address these challenges as well.

a code description like “incoming commitment” to the code itself, “11”, will help make the source data in a spreadsheet a lot easier to understand. This is not (yet) part of this research.

Approach

The prevailing approach is to tackle the challenges in this order:

- Find a way to combine all the data into a single spreadsheet.
- Use a tool to transform this single spreadsheet into IATI XML.
- Publish it yourself.

If an organisation has separate systems for parts of the information, the “merge” process is still a challenge. Also, IATI lets you fill in multiple complex values for certain fields: sector codes, names and percentages; policy markers and significance; recipient countries and regions with names, vocabularies, percentages; transactions and results with even more complex structures. Combining these hierarchical structures in a flat two-dimensional spreadsheet requires conventions and skills.

An alternative approach is to **first convert** partial data, and **then combine** these partial data sets by merging the data. You can have multiple spreadsheets as input files:

- One contains the project list with basic data such as title or description
- Another holds the transactions, exported from a financial system.
- A third contains the results, with indicators, targets, and so on. Perhaps even different files per project.

As long as they all contain a proper activity identifier, combining the data after conversion is relatively straightforward.

You could combine spreadsheets as *worksheets* in for instance an Excel or Open Document file, or you could have different spreadsheets, maybe even in different formats (Excel, CSV)³.

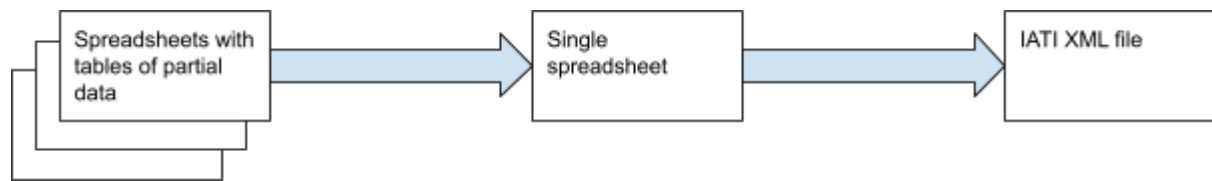
Technical architecture scenarios

“Spreadsheet Merger”

In this scenario, the individual source spreadsheets are merged into a single spreadsheet file with multiple worksheets. Part of this merge process could be to transform the data towards a standardised model.

Next, this compiled spreadsheet is converted into IATI XML.

³ In theory, the data integration could be extended to also include processing of other formats, such as JSON, SQL, or even data in R or SPSS.

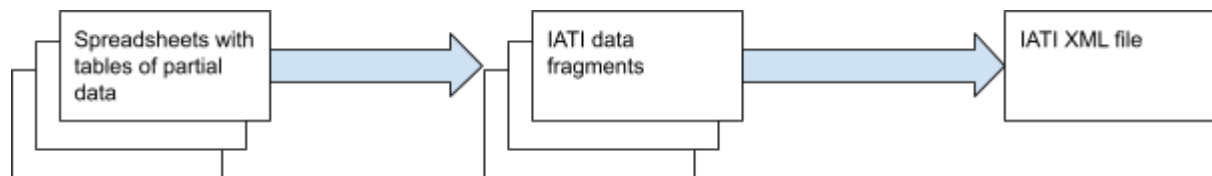


- The first step typically is done manually now, but could be automated.
- The second step could be using a “fixed model” (as is the case for the 360Giving spreadsheet conversion tool in CoVE), or could include a “data modeler” (as is the case with the CSV2IATI tool, but limited to a single sheet of data).

“Pseudo-IATI Merger”

In this scenario, the partial data is converted into “pseudo IATI”: formatted using IATI XML tags, but most likely not valid according to the IATI Schema since it is only partial data.

Next, the partial IATI XML fragments are merged, using the IATI activity identifier as the key. This produces a single IATI file.en

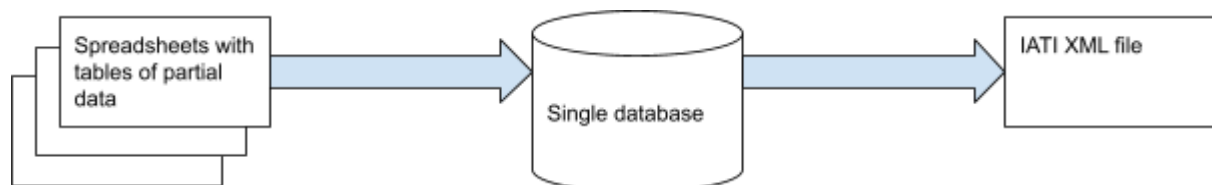


- Provenance can be recorded in a separate namespace, to allow feedback pinpointed to specific source files for data quality issues.
- A back end convert-and-merge service can be combined with various front-ends (a folder for source files, a web interface to upload spreadsheets).

“Data Integration Hub”

In this scenario, the partial data is converted and imported into a database. The database has a model of IATI data, and potentially contains more than just IATI data.

Next, the database has an IATI XML export module, to produce a single IATI XML file, or multiple files if desired.



- The database can be used by an application to provide a management interface (as is the case with AidStream and Akvo RSR) and offer presentation or visualisation interfaces (Akvo RSR and IATISudio).
- However, a typical risk is that fixes in the database are not fed back into the source systems that produced the spreadsheets.
- Existing database-based products don't seem to follow a "warehousing approach": data that no longer exists in the source files has to be manually removed from the database.

Data modeling

In all scenarios, the data in the source files needs to be modeled.

- The model can be flexible: the CSV2IATI does this.
- The model can be fixed: AidStream provides templates to fill in.

Outside IATI, two approaches are worth mentioning:

- The [360Giving Standard](#) provides templates for spreadsheets with partial data, that can be combined into a single spreadsheet and converted. The column headers in the source files have to follow certain conventions.
 - Repeatable elements (e.g. sectors, countries) are captured in separate spreadsheets.
- The [HXL Standard](#) provides a light-weight way to add HXL tags to existing spreadsheets to allow automated processing. A similar approach could be adopted: allow any column header, and add a line with tags to indicate which IATI element it represents.
 - Simple repeatable elements (single values, not compound data such as transactions) can be captured as separate columns in a spreadsheet.

A combination of these approaches can be explored.

Next steps

This first inventory and vision suggests to further look into CoVE as a possible platform for the scenario of converting a single spreadsheet with multiple worksheets with data.

Appendix

What tools are available or under development?

In no particular order.

csv2iati

Online conversion service: <http://csv2iati.iatistandard.org>

- User sets up a model to map columns to IATI fields based on a CSV file
- The same model can then applied to new CSV files in the same format

Pro	Con
• available now • offers freedom in how the data is organised in the spreadsheet • data is somewhat private	• no longer maintained • end-of-life by end 2016 or in 2017 • does IATI 1.0x (with a hack it is possible to generate 2.0x) • high learning/setup curve, complex for 'multiple value'-fields like sectors, countries • errors are hard to interpret

CoVE

Online conversion service: <http://cove.opendataservices.coop>

“Convert, Validate, Explore”: online service to translate spreadsheets into standard data. The architecture should allow to work with multiple standards, although currently it only does the 360Giving Standard.

Pro	Con
• accepts CSV, XSLX and JSON data • provides feedback specifying which spreadsheet cell contains an error • can become a service for multiple standards • under active development	• no IATI support and not planned yet • early stage product

AidStream

Online data entry service: <http://aidstream.org>

- User can enter and edit activity data in IATI format
- Import of CSV data is available for basic information about all activities, and for transactions per activity, but not for results.

Pro	Con
• reasonably user-friendly • basic validation of the standard • can be integrated with the Registry for automatic publishing	• no “dataset-level” import or conversion • needs separate CSV files of transactions per activity

• incomplete activities can be entered and remain in 'draft' status without being published • under active development	• CSV format is fixed ("simple" and "advanced" formats) • data in IATI format only available after publishing activities • organisation-oriented, cannot convert partial or multiple datasets • not intended for large data sets (more than a few dozen activities)
---	--

Akvo RSR

Online data entry service: <http://akvo.org/products/rsr/>

- User can enter and edit activity data in IATI format
- Paid service, does a lot more than produce IATI data

Pro	Con
• reasonably user-friendly • basic validation of the standard • under active development	• "dataset-level" import or conversion is not a feature but a service? • organisation-oriented, cannot convert partial or multiple datasets • is a complete "portfolio management" service, not a conversion or data entry tool

IATI Studio

Online data visualisation service: <https://www.iatistudio.com>

User has access to all published IATI data to create, save and share charts.

An IATI Publisher module is under development (expected Q3 2016), to import and convert CSV data into IATI data, and validate it according to various rule sets. Further details not published.

Pro	Con
• based on OIPA, with experience in parsing IATI data • under active development • broad range of funders/donors	• features are not public yet

Inhouse?

There are a number of organisations that are looking at ways to convert spreadsheets, based on the output of their inhouse application (Salesforce, Pluriform, Exact).

IATI Workbench

Inhouse tool set of IATI XML processing tools for data quality feedback and traceability research.

Pro	Con
• XML-based: working with XSLT and Xquery and preserving the IATI format and data in the processing • can work with any IATI data set: a complete file of an organisation, a selection of activities across all publishers • works stand-alone and offline • under active development	• not public yet • still a backend-oriented set of tools, no front-end or interface